Testing measurement invariance in the International Social Survey Program Health 2011 – the mental well-being scale

van Deurzen, I.A.; Roosma, F.

Publication date:
2014

Document Version
Other version

Testing measurement invariance in the International Social Survey Program Health 2011 – the mental well-being scale

Ioana van Deurzen[a1], Femke Roosma[a]

Introduction

The present contribution addresses the problem of measurement invariance of the mental well-being scale used by the International Social Survey Program (ISSP) Health 2011 module [1]. There are various sources of bias that can affect the comparability of a construct: some constructs are culture specific or do not exist in some countries, individual items may have a specific contextual meaning, or there can be cultural traits that affect response styles and cause method bias [2]. If the measurement invariance of a specific construct is not established then any exercise of ranking and comparing countries based on their average scores is susceptible to lead to erroneous conclusions.

The ISSP Health module was implemented in 2011 for the first time and cross-country equivalence of the scales used in this cross-national survey is unknown. The module included a battery of questions assessing individuals' health status, with items selected and adapted from established scales such as SF-8 [3] or World Health Organization health inventory [4]. Mental well-being was measured by asking respondents to evaluate how often during the past 4 weeks have they: felt unhappy and depressed (item V39); lost confidence in themselves (item V40) and felt they could not overcome their problems (item V41) [1]. The response scale ranged from 1 – "never" to 5 "very often", thus a higher score indicated worse mental well-being. Note also that all items are targeting negative aspects of well-being. For our test of measurement invariance we used the data release version 2.0.0 that was made public on the 27th August 2013 and covered 29 countries, i.e., Australia, Belgium, Bulgaria, Chile, Taiwan, Croatia, Czech Republic, Denmark, Finland, France, Germany, Israel, Japan, South Korea, Lithuania, Netherlands, Norway, Philippines, Poland, Portugal, Russia, Slovak Republic, Slovenia, South Africa, Sweden, Switzerland, Turkey, Great Britain and United States.

Measurement invariance in cross-country comparative research

In order to test measurement equivalence across countries, multi-group confirmatory factor analysis (MGCFA) is accounted as the most adequate technique [5]. In MGCFA a confirmatory factor model is tested to be equivalent across

[1] Tilburg University, Sociology Department

different groups (in our case countries), while it is possible to let different parameters of the model (slopes, intercepts, etc.) vary across groups.

There are different hierarchical levels of equivalence of measurement constructs that can be distinguished and tested. First, *configural equivalence* implies identical factor structures across countries. This test is comparable with a range of factor analyses for each country testing whether the chosen items tap into the same number of latent factors. Second, *metric equivalence* is distinguished, in which all slopes of the latent factor on the different items (i.e. factor loadings) are set equal across countries. This assures that the latent variable has the same meaning across countries. If we want to compare the means of the latent construct, we need *scalar equivalence* in which in addition to the factor structure and the factor loadings also the item intercepts are constrained to be equal across groups. Scalar invariance thus implies that differences in the means of the observed items are due to differences in the means of the underlying construct(s). To compare means of latent constructs over countries at least *partial scalar equivalence* is necessary, meaning that only two intercepts need to be constrained [5,6] .

Testing strategy

Following Meuleman & Billet [7] we started our measurement invariance tests from the highest level of measurement equivalence, scalar equivalence. In evaluating the model fit it is common to assess different goodness of fit indexes. We chose indexes that are normally used in MGCFA: Chi-Square index, the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and the Root Mean Squared Error of Approximation (RMSEA). The latter indices are normally indicated as superior to the Chi-Square index, since they take sample size and model complexity into account. In general the RMSEA is regarded to be least affected by sample size and numbers of items and factors. However, Cheung and Rensvold [8] have shown that in MGCFA the CFI, instead of the RMSEA, is the most unbiased fit statistic. The RMSEA indicates a good fit when it is < .05. An RMSEA between .05 and .08 indicates an acceptable fit. The CFI is regarded as a good fit when it is > .970, or acceptable between .950 and .970. The TFI shows a good fit when its value is > .90.

Results and discussion

Table 1 presents the results of testing measurement invariance for the ISSP Health 2011 mental well-being scale. Model 1 is the strictest model as it tests *full scalar invariance*, i.e., the three factor loadings and the intercepts of the three items were constrained to be equal across countries. In order to identify the model we constrained the loading for the item V39 to 1. For this model the RMSEA fit index showed a value far above the accepted cut-point of .08.

Despite the fact that the CFI and TLI did indicate a good fit of the model, we concluded that full scalar invariance does not apply to this particular scale.

According to the literature, a model that has a minimum of two invariant intercepts between countries (i.e., has partial scalar invariance) is sufficient for meaningful comparisons of the country mean scores [5]. We allowed in subsequent models that only one intercept at a time to vary between countries. Model 2 in Table 1 presents the fit indices for a model where we allowed the intercept of the item V39 to vary, in Model 3 in Table 1 we allowed the intercept of the item V40 to vary, while in Model 4 in Table 1 we allowed the intercept of the item V41 to vary. Model 2 in Table 1 had the best fit measures: the reduction in Chi-square was the highest and the CFI and the TLI were also in the range of the accepted values. Only RMSEA was still above the cut-point of .08.

Next we assessed the contribution of each country to the Chi-Square. A relatively high contribution to the Chi-Square could be an indication that there are specific contextual circumstances that cause construct, method or item bias. We found that South Korea stood out as an obvious outlier with a contribution to Chi-Square of 250.84 (results not presented, available from the authors). Looking at the modification indices for this country it appeared that constraining the loadings of the three items to be equal with the loadings in the other countries was a miss-specification. We concluded that in South Korea the three items do not tap into the same concept as in the other countries. Unfortunately we do not have a theoretical explanation for this. We re-estimated Model 2 without this country and we observed a substantial improvement in the fit indices: RMSEA was reduced from .099 to .090 and values for CFI and TLI also improved (Model 5 in Table 1).

Despite this strong improvement of the model fit, the value of RMSEA was still above the cut-point of .08. Looking at the countries' contribution to the Chi-Square we found three countries that could have been considered outliers: Philippine with a contribution of 103.26, Australia with a contribution of 101.78 and Great Britain with a contribution of 100.10. These countries did not stand out as strong outliers and eliminating them did not improve the model fit substantially.

From our measures of fit only RMSEA was 10% above the cut-point of .08, but this was expected due to the fact that our MGCFA model only includes three items and one latent factor resulting in a RMSEA with large standard errors and a biased fit score [8]. We decided to accept Model 5 in Table 1 as our final model, which implies that the ISSP Health 2011 mental well-being scale is partial scalar invariant in the sample of 28 countries, after excluding South Korea.

In Table 2 we present the country rankings based on the average latent factor scores. The Spearman correlation coefficient between the rankings based on the latent and the observed scores was .95 ($p<.000$) indicating that differences in the observed factor scores are due to differences in the underlying concept. We then compiled information on the ranking based on other measures of mental well-being, namely the restricted Center for Epidemiologic Studies Depression Scale (CES-D 8) [9] and the World Health Organization positive well-being scale (WHO-5) [10] for European countries that are also present in the ISSP Health 2011 sample. The Spearman correlation coefficient between the rankings on the three mental well-being scales is summarized in Table 3 and they point to the fact that the ISSP mental well-being scale taps into a different dimension that the CES-D 8 and WHO-5.

In conclusion, even if the ISSP mental well-being scale is partial scalar invariant in 28 countries it is remarkable that its content is not comparable with related scales of mental well-being. We encourage scholars to examine the cause of this in further research.

References

1 ISSP Research Group, International Social Survey Programme (ISSP): Health and Health Care (2011). Distributor: GESIS Cologne Germany ZA5800, Data Version v2.0.0 (2013-08-27).

2 Van de Vijver, FJR. (1998). Towards a theory of bias and equivalence. Zuma Nachrichten, 3, 41-65.

3 Ware J, Kosinski M, Dewey J, Gandek B (2001). How to Score and Interpret Single-Item Health Status Measures: A Manual for Users of the SF-8 Health Survey.Boston: QualyMetric.

4 Bech, P. (2004). Measuring the Dimension of Psychological General Well-Being by the WHO-5. Quality of Life

5 Steenkamp, J.B.E.M., & Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. Journal of Consumer Research, 25(1), 78-107.

6 Davidov, E. (2008). A Cross-Country and Cross-Time Comparison of the Human Values Measurements with the Second Round of the European Social Survey. Survey Research Methods, 2(1), 33-46.

7 Meuleman, B. & Billet, J. (2011). Religious Involvement: Its Relation to Values and Social Attitudes. In E. Davidov, P. Schmidt & J. Billet (Ed.), Cross-Cultural Analysis.

8 Cheung, G.W., & Rensvold, R.B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. Structural Equation Modeling: A Multidisciplinary Journal, 9(2), 233-255.

9 van de Velde, S., Bracke, P., Levecque, K., Meuleman, B. (2010). Gender differences in depression in 25 European countries after eliminating measurement bias in the CES-D 8. Social Science Research, 39, 396-404.

10 European Foundation for the Improvement of Living and Working Conditions, European Quality of Life Survey, 2011-2012 [computer file]. Colchester, Essex: UK Data Archive [distributor], July 2013. SN: 7316.

Table 1: Model specifications and fit indices

Model		Chi-Square (df)	Df	Δ Chi square (Model – Model1)	RMSEA	CFI	TLI	N groups
Full Scalar Invariance								
1	Intercepts and loadings equivalent	2689.639 (112)	112	-	0.125	0.961	0.969	29
Partial Scalar Invariance								
2	Intercept item c) varies across countries	**1302.088 (84)**	**84**	**1387.551**	**0.099**	**0.981**	**0.981**	**29**
3	Intercept item d) varies across countries	1667.612 (84)	84	1022.027	0.113	0.976	0.975	29
4	Intercept item e) varies across countries	1868.525 (84)	84	821.114	0.120	0.973	0.972	29
Partical Scalar Invariance (adjusted number groups)								
5	***Model 2 without South Korea***	***1043.038 (81)***	***81***	***1646.601***	***0.090***	***0.985***	***0.984***	***28***

Table 2. Rankings based on the ISSS Health 2011 mental well-being scale (rank 1 means worse mental well-being, rank 28 means best mental well-being)

Country	Rank based on the latent factor scores
Russia	1
Japan	2
Belgium	3
France	4
Turkey	5
Great Britain	6
Australia	7
Portugal	8
Lithuania	9
Chile	10
Sweden	11
Bulgaria	12
Norway	13
Croatia	14
Finland	15
Netherlands	16
South Africa	17
Denmark	18
Germany	19
Philippine	20
Slovakia	21
Israel	22
Taiwan	23

Czech Republic	24
Poland	25
United States	26
Slovenia	27
Switzerland	28

Table 3. Spearman rank correlation coefficients between the rankings of the countries based on the ISSP Health 2011 mental well-being scale, CES-D 8 (15 countries) and WHO-5 scale (16 countries)

Ranking based on:	CES-D 8	WHO-5
ISSP Health 2011	.303 (p=.25)	.11 (p=.67)
CES-D 8	1	**.65** (p=.02)